CLAIMS:

What is claimed is:

1. A method, in a data processing system, of distributing traffic to application instances on one or more computing devices, comprising:

obtaining application instance specific operational information identifying operational characteristics of an application instance on a computing device of the one or more computing devices;

generating a load balancing weight to be associated with the application instance based on the application instance specific operation information obtained; and

distributing traffic to the application instance based on the generated load balancing weight.

2. The method of claim 1, wherein obtaining application instance specific operational information includes:

retrieving the application instance specific operational information from the application instance using an agent application resident on the computing device.

3. The method of claim 2, wherein the application instance is instrumented to include code for communicating with the agent application and sending application instance specific operational information to the agent application from the application instance.

4. The method of claim 1, wherein generating a load balancing weight to be associated with the application instance includes:

comparing the application instance specific information to one or more other application instance specific information for one or more other application instances; and

generating a load balancing weight based on a relationship between the application instance specific information and the other application instance specific information.

5. The method of claim 4, wherein the relationship is a relative difference between the application instance specific information and the one or more other application instance specific information.

6. The method of claim 4, wherein generating a load balancing weight based on a relationship between the application instance specific information and the other application instance specific information includes:

attributing weight points to each application instance of the application instance and the one or more other application instances based on a relative difference between the application instance specific information associated with the application instance and the other application instance specific information associated with the one or more other application instances.

7. The method of claim 1, wherein the application instance specific information includes at least one of a number of successful transaction processed by the application instance within a period of time, an application instance response time, an application instance topology, an importance of transactions currently being processed by the application instance, an amount of time the application instance has been blocked waiting for resources, and an amount of resources consumed by the application instance.

8. The method of claim 2, wherein retrieving the application instance specific information from the agent application is performed periodically.

9. The method of claim 1, wherein the method is implemented in a weight management system that is separate from the computing devices and from a load balancing device.

10. The method of claim 4, wherein generating a load balancing weight based on a relationship between the application instance specific information and the other application instance specific information includes:

assigning a base weight to each of the application instance and the one or more other application instances; and

increasing a weight value associated with the application instance or an other application instance based on one or more of the following:

which application instance has a relatively higher transaction success rate;

which application instance has a relatively better response time;

which application instance operates on an underutilized system;

which application instance has a relatively better response time and operates on an underutilized system;

which application instance processes the least significant transactions; and

which application instance passes transactions on to higher performing computing systems.

11. A computer program product in a computer readable medium for distributing traffic to application instances on one or more computing devices, comprising:

first instructions for obtaining application instance specific operational

information identifying operational characteristics of an application instance on a computing device of the one or more computing devices;

second instructions for generating a load balancing weight to be associated with the application instance based on the application instance specific operation information obtained; and

third instructions for distributing traffic to the application instance based on the generated load balancing weight.

12. The computer program product of claim 11, wherein the first instructions for obtaining application instance specific operational information include:

instructions for retrieving the application instance specific operational information from the application instance using an agent application resident on the computing device.

13. The computer program product of claim 12, wherein the application instance is instrumented to include code for communicating with the agent application and sending application instance specific operational information to the agent application from the application instance.

14. The computer program product of claim 11, wherein the second instructions for generating a load balancing weight to be associated with the application instance include:

instructions for comparing the application instance specific information to one or more other application instance specific information for one or more other application instances; and

instructions for generating a load balancing weight based on a relationship

between the application instance specific information and the other application instance specific information.

15. The computer program product of claim 14, wherein the relationship is a relative difference between the application instance specific information and the one or more other application instance specific information.

16. The computer program product of claim 14, wherein the instructions for generating a load balancing weight based on a relationship between the application instance specific information and the other application instance specific information include:

instructions for attributing weight points to each application instance of the application instance and the one or more other application instances based on a relative difference between the application instance specific information associated with the application instance and the other application instance specific information associated with the one or more other application instances.

17. The computer program product of claim 11, wherein the application instance specific information includes at least one of a number of successful transaction processed by the application instance within a period of time, an application instance response time, an application instance topology, an importance of transactions currently being processed by the application instance, an amount of time the application instance has been blocked waiting for resources, and an amount of resources consumed by the application instance.

18. The computer program product of claim 12, wherein the instructions for retrieving the application instance specific information from the agent application is performed periodically.

19. The computer program product of claim 14, wherein the instructions for generating a load balancing weight based on a relationship between the application instance specific information and the other application instance specific information include:

instructions for assigning a base weight to each of the application instance and the one or more other application instances; and

instructions for increasing a weight value associated with the application instance or an other application instance based on one or more of the following:

which application instance has a relatively higher transaction success rate;

which application instance has a relatively better response time;

which application instance operates on an underutilized system;

which application instance has a relatively better response time and operates on an underutilized system;

which application instance processes the least significant transactions; and

which application instance passes transactions on to higher performing computing systems.

20. A system for distributing traffic to application instances on one or more computing devices, comprising:

means for obtaining application instance specific operational information identifying operational characteristics of an application instance on a computing device of the one or more computing devices;

means for generating a load balancing weight to be associated with the application instance based on the application instance specific operation information obtained; and

means for distributing traffic to the application instance based on the generated load balancing weight.